THE OREGON STATE DRUG REVIEW©
AN EVIDENCE BASED DRUG THERAPY RESOURCE http://pharmacy.oregonstate.edu/drug-policy/newsletter
December Volume 15 Issue 10

Evaluating the Quality of Evidence from Observational Studies: Lessons from Acetaminophen Use in Pregnancy

Kathy Sentena, PharmD, Oregon State University Drug Use Research and Management Group and Rose Hong, Pharm D Candidate, Oregon State University

Observational studies are a valuable component of medical evidence, particularly when randomized controlled trials (RCTs) are not feasible. In fields such as obstetrics, where studies involving pregnant women would be unethical, much of what we know about medication safety arises from observational data. However, observational study conclusions depend heavily on how well the investigators control for study design factors, and even the best designed observation studies carry inherent bias.

The recent attention on the safety of acetaminophen (APAP) use during pregnancy provides a valuable example for examining these issues. There are numerous observational cohort and case-control studies investigating possible links between maternal APAP use and childhood neurodevelopmental disorders (NDDs), such as attention-deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD). Some studies found a possible association, while others found no association. The discrepancies between them illustrate the importance of understanding the limitations of what can be inferred by observational studies.

This newsletter describes key factors to determining the quality of observational evidence, such as study design, confounding control, exposure assessment (i.e., documentation of treatment administration), and outcome measurement to assist readers in critically evaluating observational studies.

Background

Observational studies, also called epidemiological studies, are when researchers observe the effect of a treatment or other intervention without influencing the exposure to treatment.[1] Different types of observational study designs and their characteristics are described in **Table 1**.

Table 1. Observational Study Designs*[1,2]

Study Design	Measures of Association	Design Advantages	Design Disadvantages
Case Report or case series	Unable to calculate	Identify new disease or adverse events and hypothesis generation	Unable to calculate rates , risk or measure association
Cross-sectional	Point prevalence or odds ratio	Quick and inexpensive	Temporal association cannot be established
Case-control	Odds ratio	Study of rare outcomes and less expensive and faster than cohort	Not practical for study of rare exposure to treatment
Cohort	Relative risk, odds ratio or hazard ratio	Study of rare exposures and temporal association most clearly delineated. Direct estimates of risk and incidence can be calculated High risk of recall bias	Not as practical for study of rare outcomes. Increased cost and length

***Data are collected retrospectively except for cohort studies which can collect data prospectively and retrospectively.**

Retrospective studies are those where data are collected from the past, either through medical records or by asking participants to recall their treatments or outcomes.[3] Retrospective studies cannot demonstrate temporality (e.g., relating to the sequence of time or to a particular time) as accurately and are more prone to several biases, particularly recall bias.[1] However, with the widespread use of electronic health records (EHR) the use of existing electronic data is more common, which minimizes the concern of temporality in case-control studies. Temporality is required to establish causation; however, temporality alone is not sufficient to demonstrate causality. Prospective studies recruit and follow participants forward through time, collecting data in the process. Prospective studies are less prone to some types of biases and can more easily demonstrate that the treatment preceded an outcome, thereby more strongly suggesting causation.[1]

1. Study Design and Cohort Selection

The strength of the conclusion derived from an observational study is dependent upon its design and methods.[3] A cohort selection follows a specific group of people based on a shared characteristic.[1] Cohort studies can be prospective and follow participants over time to assess outcomes after treatment exposure, while case-control studies are always retrospective and compare exposure history between those with and without an outcome.[2] How data are collected, how participants are

selected, and how the data are analyzed all influence study validity.

Determining the characteristics of the population studied is an important component of study design. Clearly defined eligibility criteria that is representative of the target population reduces variability (i.e., heterogeneity) so that stronger conclusions can be drawn. Many observational studies of maternal APAP use failed to produce reliable results because participants taking APAP also had higher rates of infections, depression, and chronic pain compared to non-users. It is important to note that observational studies allow for extrapolation of findings to a more general population which may be excluded from traditional RCTs.

Example: The Danish National Birth Cohort study followed 64,322 women during pregnancy and 6 months after pregnancy from 1996-2002 to research if there was an association with APAP use with fetal brain development.[4] Eligibility criteria was clearly defined enrolling only those with single births, complete records of APAP use, and documentation of measured major confounders. The study evaluated exposure rates in a broad general population, increasing applicability.

Takeaway: High quality observational studies are those that use large, prospective analysis of patients with similar characteristics (e.g., well designed eligibility criteria) to decrease bias and enhance the credibility of the results.

2. Implications of Recall Bias

Recall bias is under a larger umbrella of "information bias" which focuses on the accuracy with which data are collected, how they are interpreted and recorded, and how they are identified by the researchers. Study participants may be asked to recall medication timing, dosing and other administration details in retrospective observational studies, such as case-control designs. This type of design lends itself to recall bias, as participants often fail to remember specifics around therapies that were taken in the past.[3] Participants who had a negative outcome, such as a child developing autism, may be more likely to remember details around medication administration and even overreport use compared to those who did not have a negative experience.

Example: A 2021 study was conducted in a population-based cohort of more than 73,000 mother-child pairs across six European countries.[5] APAP use was determined through maternal questionnaires or interviews conducted prenatally and up to 18 months after birth.

The large sample size and multicenter design strengthen applicability; however, recall bias is a limitation in this example. Relying on self-reporting of exposure months or even years earlier can bias findings towards mothers who had children with developmental issues. Recall bias may influence results by magnifying apparent associations.

This study did minimize recall bias by collecting

Takeaway: When evaluating the effect of recall bias in an observational study, consider whether the exposure data were collected prospectively and assess the timeframe of recall that was utilized in studies with a retrospective design. Assess if exposure information is drawn from surveys or medical records as well as the completeness of medical records for the exposure under investigation.

3.The Effect of Confounding

A confounder is a variable associated with both the exposure and the outcome, which may create a distorted relationship.[3] Observational trials lack randomization, which can cause groups to differ for known as well as unknown factors, making cause and effect difficult to determine and subject to confounding. For example, in the studies of APAP use in pregnancy, the presence of a fever may increase the chance a mother takes acetaminophen. The fever itself may increase the risk of an NDD. Confounding can lead to false associations or mask real ones.[3] Statistical adjustments like multivariable regression help but rarely eliminate all confounding, especially for unmeasured variables.

Example: The US Nurses' Health Study II minimized confounding by employing a negative control exposure (NCE) design.[6] Negative control exposure evaluates residual confounding in observational studies, which is a treatment that is not causally related to the outcome but is affected by the same confounding factor as the exposure of interest (APAP).[6] This study prospectively evaluated APAP use in mothers that had children (n=8,856) born between 1993-2005.[6] This approach compared mothers who used APAP during pregnancy to those who used it before or after pregnancy. The NCE method helps to detect whether observed associations might be due to shared maternal factors rather than the exposure itself. The study found that in-pregnancy exposure to APAP, but not pre- or post-pregnancy use APAP, was modestly associated with ADHD (hazard ratio [HR] 1.34; 95% CI, 1.05–1.72).[6]

An additional example of limiting unmeasured cofounding is illustrated in a sibling control analysis. A sibling control analysis provides higher quality observational data that adjust for potential confounders (genetics, maternal health factors, acetaminophen dosage, etc.) by holding the mother constant.

Example: A nationwide cohort study in Sweden utilized a sibling control.[7] The study evaluated the association of APAP use in pregnancy and risk of ASD, ADHD, and intellectual disability.[7] The study concluded that APAP was not associated with ASD, ADHD, or intellectual disability (HR 0.85 to 0.88; p>0.05 for low, medium and high dose comparisons).[7]

Takeaway: Controlling for confounders is an important part of observational research. Uncontrolled confounders that are not measured or adjusted for can result in conclusions that are not reliable. Even observational data that controls for confounders can result in contradictory results as illustrated between the US Nurses' Health Study II and the cohort study in Sweden.

4. Exposure Assessment: Accuracy and Timing of Measurement

Another important component of observational research is how the exposure to treatment is measured. Details on how the study precisely captures when, how much, and for how long participants were exposed to the medication being studied can influence the conclusions.

Example: In the Boston Birth Cohort study, plasma biomarkers of APAP were measured in 996 mother–infant pairs between 1998 and 2018.[8] Because exposure was confirmed by biologic samples rather than maternal recall, this design reduces recall bias substantially. The researchers found a weak association between higher biomarker levels and an increased risk of developmental diagnoses such as ADHD and ASD.[8]

Takeaway: When evaluating an observational study determine if medication use was self-reported or objectively measured, how often it was assessed, and whether timing and dosage were recorded with sufficient accuracy to support conclusions.

However, even biomarker studies have limitations. A single plasma measurement represents only recent or short-term use, not the full exposure window during pregnancy. Therefore, while exposure misclassification is lower, temporal precision (which trimester, how often, at what dosage) remains imperfect. Confounders may still be present complicating interpretation of findings.

5. Outcome Measurement and Diagnostic Validity

A major determinant of study quality is how outcomes are defined and verified. Developmental disorders such as ADHD and ASD are behaviorally defined disorders that rely on diagnostic criteria which has changed over time and may vary between studies and provider assessments.

Example: The Danish National Birth Cohort followed children for seven years and assessed ADHD using parental reports, medication data, and clinical diagnoses.[4] While this strengthens outcome verification, it still introduces variability. Some parents may seek diagnoses more readily, or clinicians may differ in diagnostic thresholds.

In contrast, the evaluation of the US Nurses' Health Study II used validated diagnostic instruments and a NCE analysis to strengthen confidence in ADHD and ASD classification.[6] Quality of the data is still subject to accurate medical record linkage, consistent diagnosis coding criteria, and similar follow-up durations across groups.

Takeaway: Outcomes need to be clearly defined; validated diagnostic tools should be used and consistent follow-up time frames applied to avoid selection bias.

Additional Considerations

Beyond study design and cohort selection, several other key factors must be considered when evaluating the quality of observational studies. These factors determine whether findings are valid, reproducible, and potentially associated. **Table 2** highlights other potential limitations that should be considered. Limitations of observational studies can impact the study conclusions to a small or large extent.

Additionally, single studies should rarely be relied on to draw strong conclusions. Critical appraisal of the literature, regardless of study type, is needed to determine the degree of bias. Evaluating observational studies requires assessment of strengths and weaknesses to determine the causality. Adding what is already known from other studies and combining it with new data on a topic is important in clinical decision making. Observational studies provide value by contributing in this way.

Table 2. Limitations of Observational Studies[3]

Limitation	Explanation
Selection Bias	Occurs when participants selected for the study are different from those not included in ways related to exposure to the treatment and outcome.
Reverse Causation	Unclear if the exposure to treatment caused the outcome or the outcome influenced the exposure.
Residual Confounding	Unmeasured or poorly measured variables can distort results even after adjusting for measured confounders.
Observer and Reporting Bias	Investigators or participants may influence data collection or reporting.
Publication Bias	Studies with positive or significant results are more likely to be published than null findings.

Conclusion

Observational studies vary by design and methods, and these differences influence their inherent biases, confounding, and validity. Observational studies can identify potential associations but cannot establish causality. A body of evidence is needed, despite the study design, to determine causality. Additionally, associations between an exposure and an outcome should not be concluded by one study alone; multiple studies across different populations should be considered when making such assessments.[1] Use shared decision making with patients after discussing potential benefits and risks of therapies, especially when only observational evidence is available.

Peer Reviewer: Jessina C. McGregor, PhD FSHEA, Professor, Department of Pharmacy Practice, College of Pharmacy, Oregon State University

References:

1. Thiese MS. Observational and interventional study design types; an overview. *Biochem Med (Zagreb)*. 2014;24(2):199-210. doi:10.11613/BM.2014.022

2. DiPietro NA. Methods in epidemiology: observational study designs. *Pharmacotherapy*. 2010;30(10):973-984. doi:10.1592/phco.30.10.973

3. Hannan EL. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC Cardiovasc Interv*. 2008;1(3):211-217. doi:10.1016/j.jcin.2008.01.008

4. Liew Z, Ritz B, Rebordosa C. Acetaminophen use during pregnancy, behavioral problems, and hyperkinetic disorders. *JAMA Pediatics.* 2014;168(4):313-320.

5. Alemany S, Avella-Garcia C, Liew Z, et al. Prenatal and postnatal exposure to acetaminophen in relation to autism spectrum and attention-deficit and hyperactivity symptoms in childhood: meta-analysis in six European population-based cohorts. *Eur J Epidemiol.* 2021;36(10):993-1004.

6. Liew Z, Kioumourtzoglou MA, Roberts AL, O'Reilly ÉJ, Ascherio A, Weisskopf MG. Use of Negative Control Exposure Analysis to Evaluate Confounding: An Example of Acetaminophen Exposure and Attention-Deficit/Hyperactivity Disorder in Nurses' Health Study II. *Am J Epidemiol*. 2019;188(4):768-775. doi:10.1093/aje/kwy288

7. Ahlqvist, V.H., Sjöqvist, H., Dalman, C., Karlsson, H., Stephansson, O., Johansson, S., Magnusson, C., Gardner, R.M., Lee, B.K., 2024. Acetaminophen Use During Pregnancy and Children's Risk of Autism, ADHD, and Intellectual Disability. *JAMA*. 2025;331(14):1205-1214.

8. Ji Y, Azuine RE, Zhang Y, et al. Association of Cord Plasma Biomarkers of In Utero Acetaminophen Exposure With Risk of Attention-Deficit/Hyperactivity Disorder and Autism Spectrum Disorder in Childhood. *JAMA Psychiatry*. 2020;77(2):180-189. doi:10.1001/jamapsychiatry.2019.3259